片上 TCAM 的研究和应用

许 俊[1,2]，龚源泉[1]，李 丽[2]
（1.盛科网络苏州有限公司，江苏 苏州 215021；
2.南京大学 江苏省光电信息功能材料重点实验室，江苏 南京 210093）

摘 要：为了提高片上 TCAM 的摆放密度和降低功耗，基于 IBM 32 nm 工艺库提供的 TCAM 的特性和优先编码器硬核，设计出同时满足多个查找宽度的外围控制电路。相比于之前的设计和实现，该设计可以减少 TCAM 的块数和相关寄存器的数量，减少片上 TCAM 的摆放面积，降低芯片的整体功耗。该设计已经成功应用于公司第 4 代路由交换 ASIC 芯片上。

关键词：片上 TCAM；优先级编码器；ASIC
中图分类号：TN431.2 **文献标识码**：A **文章编号**：1674-7720(2013)19-0069-03

Research and application of on-chip TCAM

Xu Jun[1,2], Gong Yuanquan[1], Li Li[2]
(1.Centec Networks Suzhou Inc., Suzhou 215021, China;
2.Key Laboratory of Advanced Photonics and Electronics Materials, Nanjing University, Nanjing 210093, China)

Abstract: In order to decrease the floorplan area and lower the power consumption of TCAM on chip, the implementation is presented to fulfill multiple lookup key size at the same time with the TCAM priority and encoder hardcore which are provided by IBM 32 nm technology library. Comparing to the former implementations, the new design improves the floorplan density of TCAM, decreases the registers number and lower the power consumption of chip. The design has been used in 4th generation router switch ASIC of company successfully.

Key words: on-chip TCAM; priority encoder; ASIC

1 TCAM 简介

三态内容寻址存储器（TCAM）是一类特殊的存储器，传统的存储器都是根据地址读出内容，例如静态存储器（SRAM）和动态存取器（DRAM），但是 TCAM 是根据存储的内容得到对应的地址，输入一个数据（称为查找内容或者查找 key），TCAM 内部就把这个 key 和它所有存储的条目作并行比较，然后把匹配的地址输出，如果有多个条目都与这个查找 key 匹配，那么输出最小的地址。

TCAM 又有外挂和片上（内嵌）之分，外挂 TCAM 原来有多家厂商提供，经过一系列的并购，目前只有博通公司和瑞萨电子可以提供商用的 TCAM 芯片。外挂 TCAM 一般适用于交换容量在 100 Gb/s 量级的路由交换芯片或者安全芯片。对于带宽超过 100 Gb/s 以上的路由交换芯片或者安全芯片而言，直接在芯片内部集成 TCAM 不失为一种好的选择，特别是做并行访问控制列表（ACL）查找时，就特别需要片上 TCAM。

片上 TCAM 有多个厂家可以提供不同的工艺库，其中 IBM 的片上 TCAM 工艺库是目前为止面积最优、功耗最小且速度最快的片上 TCAM 工艺库之一。

TCAM 的基本单元由一个数据位（data）和一个掩码位（mask）构成，所以顾名思义称为三态存储器，当输入 1 bit 数据（input）时，当 input=data & mask，才算匹配。这时，TCAM 会输出一个命中（hit）指示，表示这个条目命中，这个特性让 TCAM 在 ACL、路由查表的最长前缀匹配和模糊查找中特别有用。

但是 TCAM 也有不足之处，主要体现在两个方面：一个是相比较于 SRAM 和 DRAM，它的存储密度很低，摆放密度也低；另外，TCAM 做查找的时候，功耗特别大（因为需要所有的条目并行作比较）。

图 1 显示 IBM 45 nm 工艺下的 TCAM 与 SRAM、

DRAM 的比较，TCAM 只有 1.10%的存储容量（百万比特数），却占用所有存储器 8.64%的面积（如果考虑到摆放面积，这个数字还要加倍）和消耗 10.31%的功耗。

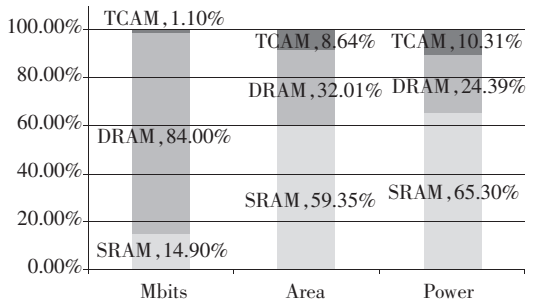


图 1 TCAM 和 SRAM、DRAM 的比较（IBM 45 nm）

本文基于 IBM 32 nm 工艺库提供的 TCAM 的一些新特性和提供的优先级编码器硬核，通过特别设计的电路，可以同时满足 160 bit、320 bit 和 640 bit 数据的查找，并且减少 TCAM 的块数、降低 TCAM 的功耗。

2 片上 TCAM 外围电路的设计

2.1 多种查找宽度电路设计

此前为了实现多种位宽 key 的查找，必须以最小 key 的位宽作为 TCAM 的位宽选择。例如，需要同时支持 160 bit、320 bit 和 640 bit 3 种 key 位宽的查找，需要选择 TCAM 的位宽必须是 160 bit 的，然后通过横向拼接的方式实现 320 bit 和 640 bit 位宽的 key 宽度查找。

IBM 32 nm 工艺库提供的 TCAM 支持一种称为列使能的方式。以 1 024 深度 320 bit 宽度的 TCAM 为例，图 2 表示将 320 bit 宽度分成 4 列，每一列有 80 bit 位宽，4 列分别是 FE0、FE1、FE2 和 FE3，其中 FE0 和 FE2 是偶数列，FE1 和 FE3 是奇数列。当进行 160 bit 查找时，FE0 和 FE2 自动拼接成 160 bit 与 key 做匹配，并且结果输出到 MLLA[1 023：0]，每个 bit 代表一个条目的查找结果，FE1 和 FE3 自动拼接成另外一个 160 bit 与 key 做匹配，并且输出结果到 MLLB[1 023：0]，通过外部设计的电路，先把 MLLA[1 023：0]中最小的匹配地址找出来（通过优先级编码器），同时把 MLLB[1 023：0]中最小匹配地址找出来，最后比较 MLLA 和 MLLB 中哪个匹配地址最小，取小优先，如果值相等，则优先取 MLLA 的结果作为最终结果。

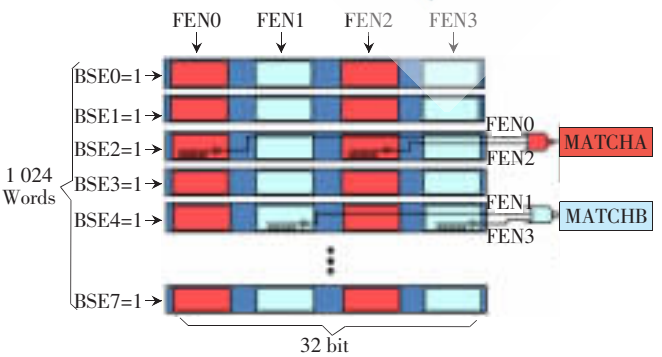


图 2 TCAM 列查找示意（IBM 32 nm）

这样，TCAM 的最小宽度就不必是 160 bit，可以是 320 bit，与之前的需要 160 bit 宽度的 TCAM 相比，构建相同大小的查找表，TCAM 的块数可以减少一半。

TCAM 的块数多少直接影响到芯片的面积大小，前面提到过 TCAM 本身物理的面积就比 DRAM 和 SRAM 大，此外，TCAM 由于在做查找时需要把输入的内容和存储的所有条目同时作比较，会导致片上供给 TCAM 的电源噪声变大。为了解决这个问题，一般需要在 TCAM 之间插入大的片上去耦电容，再加上需要把优先级编码电路和相关的寄存器紧靠着 TCAM 摆放，因此需要 TCAM 块与块之间有一定的间隔。图 3 显示了在 45 nm 工艺下 16 块 TCAM 在硅片上的摆放面积是 3.231 5 mm^2（=2.81 mm×1.15 mm），相比较于 TCAM 本身的面积（1.59 mm^2），大了将近一倍。所以，从这个层面上而言，构建相同大小的查找表，TCAM 的块数越多越不好。

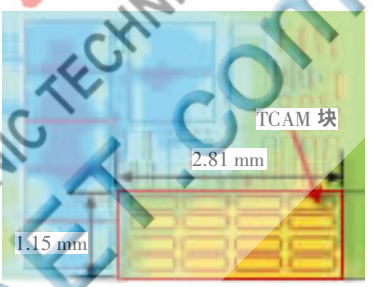


图 3 片上 TCAM 摆放面积（2.81 mm×1.15 mm，IBM 45 nm）

此前为了解决 TCAM 查找结果 MLL（匹配位）输出的时序问题，可行的做法是将匹配的结果先用寄存器锁存起来，再送给后续的优先级编码器。IBM 32 nm 的工艺库新提供了 6~64 的优先级编码器硬核，可以解决时序问题。

图 4 显示 IBM 32 nm 提供的 6~64 的优先级编码器硬核，输出信号 HIT 表示所有 MLL[63：0]作位“或”运算之后的结果。

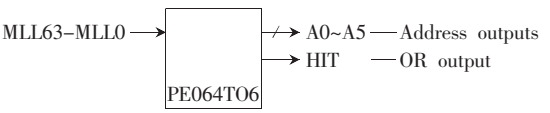
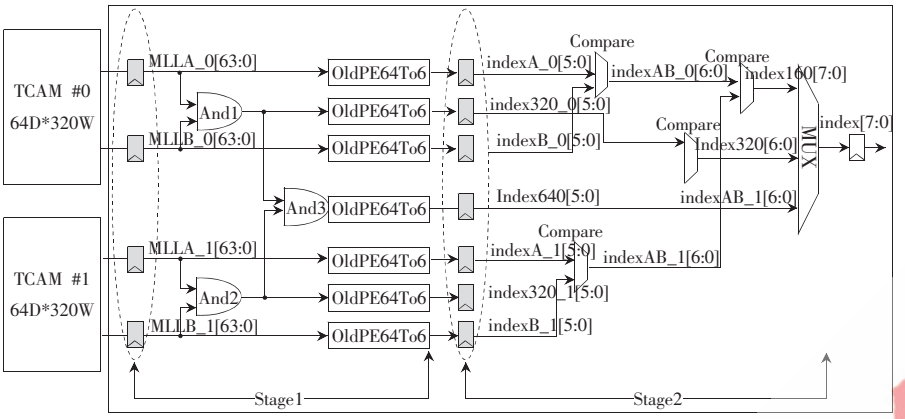


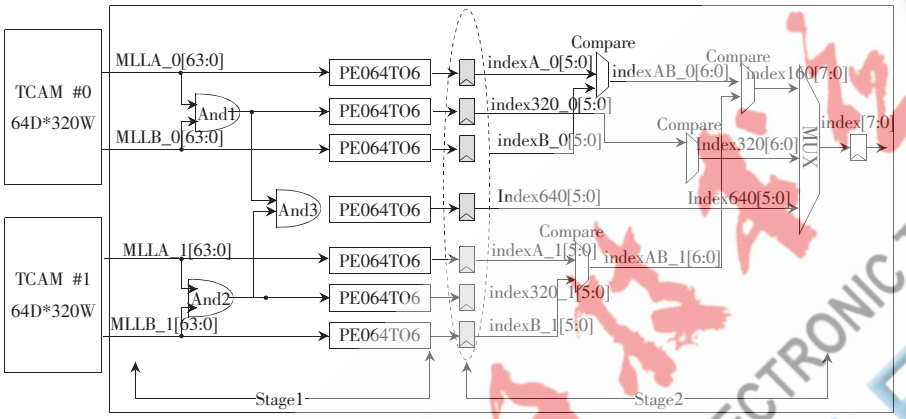
图 4 IBM 32 nm 优先级编码器硬核（IBM 32 nm）

图 5（a）中显示了采用用户自己设计的 6~64 的优先级编码器时，必须将 TCAM 输出的匹配结果先用寄存器锁存一拍，然后才送给优先级编码器。图 5（b）显示，采用 IBM 32 nm 提供的优先级编码器硬核，可以直接把 TCAM 的匹配结果经过一个或者两个与门之后与优先级编码器的输入对接，优先级编码器的输出再进入寄存器锁存起来。

表 1 比较了两种方案所需要的资源，可以看到，TCAM 的匹配结果需要的锁存寄存器可以全部节约下来，只要 TCAM 的深度越大，节约的寄存器就越多，以支



(a)此前的优先级编码器硬核以及相关电路



(b)采用优先级编码器硬核以及相关电路

图 5 优先级编码器硬核及相关电路比较

表 1 两种方案所需要的资源比较

实现方案	寄存器个数	两输入与门个数	优先级编码器个数	比较器个数	3 选 1 多路器个数
(a)	313	192	7	4	9
(b)	57	192	7	4	9

持 16 384 个 160 bit 宽度的 ACL 条目为例，方案(a)需要额外多出 16 384 个寄存器。

此外，方案(a)比方案(b)会多一级流水线的延迟。

采用图 5(b)所示的电路，可以实现 160 bit、320 bit 和 640 bit 3 种 key 宽度的查找。

160 bit 宽度 key 的查找流程如下：

(1)TCAM#0 输出匹配结果 MLLA_0 [63:0] 和 MLLB_0 [63:0]，同时，TCAM#1 输出 MLLA_1 [63:0]和 MLLB_1[63:0]。

(2)MLLA_0[63:0]经过 6~64 优先级编码器，输出 indexA_0[5:0]和 hitA_0(图中没有标示出来)。同样地，对于 MLLB_0[63:0]、MLLA_1[63:0]和 MLLB_1[63:0]经过各自对应的优先级编码器，输出 indexB_0[5:0]、indexA_1 [5:0] 和 indexB_1 [5:0]，以及对应的 hitB_0、hitA_1 和 hitB_1。

(3)indexA_0[6:0]和 indexB_0[6:0]比较，如果 hitA_0 和 hitB_0 二者只有一个为 1，那么选择对应的 index 输出；如果 hitA_0 和 hitB_0 均为 1(表示都有匹配到)，则选择 indexA_0 输出。当选中 indexA_0 时，输出 indexAB_0[6:0]={indexA_0[5:0],1'b0}，最低位补 0；当选中 indexB_0 时，输出 indexAB_0 [6:0] ={indexB_0 [5:0],1'b1}，最低位补 1，此外还需要把 hitA_0 和 hitB_0 作位“或”运算输出 hitAB_0。

(4) 对于 indexA_1 [6:0] 和 indexB_1[6:0]有同样的操作，得到结果 indexAB_1[6:0]和 hitAB_1。

(5) 比较 indexAB_0 [6:0] 和 indexAB_1[6:0]，操作过程类似于步骤(3)，最后得到 index160 [7:0] 和 hit160。

320 bit 宽度 key 的查找流程如下：

(1)TCAM#0 输出匹配结果 MLLA_0 [63:0]和 MLLB_0 [63:0]，同时，TCAM#1 输出 MLLA_1 [63:0]和 MLLB_1[63:0]。

(2)MLLA_0 [63:0] 每个比特和 MLLB_0[63:0]的每个对应比特位进行“与”运算，得到 MLLAB_0[63:0]，再输入到一个专门的 6~64 优先级编码器，输出 index320_0[5:0]和 hit320_0(图中没有标示)；对于 MLLA_1 [63:0] 和 MLLB_1[63:0]，有同样的操作，把 MLLAB_1[63:0](=MLLA_1[63:0] & MLLB_1[63:0])输出 index320_1[5:0]和 hit320_1。

(3)比较 index320_0[5:0]和 index320_1[5:0]，过程与前述类似，得到 index320[6:0]和 hit320。

640 bit 宽度 key 的查找流程如下。

(1)将前述 320 bit 宽度 key 查找流程的步骤(2)得到的 MLLAB_0[63:0]和 MLLAB_1[63:0]再作一级按位“与”操作(MLLAB_0[63:0]& MLLAB_1[63:0])，结果输出到优先级编码器中，得到 index640[5:0]和 hit640。

最后还有一级多路选择器，根据全局配置，在 index160[7:0]、index320[6:0]和 index640[5:0]三者之间选择一个作为最终结果输出

上面设计的 TCAM 查找电路，与之前的设计相比，在需要同样大小的查找表情况下，TCAM 的块数少一半，而且由于应用了优先级编码器硬核，可以把第一级的锁存寄存器全部省掉，此外还降低了 TCAM 的摆放面积和功耗。

2.2 利用 TCAM 的预查找功能降低查找功耗

IBM 32 nm 工艺库中的 TCAM 为了防止查找时的瞬间功耗过大，提供了一种预查找功能。TCAM 横向的块

称为一个 Bank，每个 Bank 包含 128 个条目，每个条目无论多少位宽，可以按照 80 bit 来切分，每 80 bit 的存储数据可以分为两级进行查找，第一级查找称为预查找，只匹配低 bit0~bit7 总共 8 bit，如果这 8 bit 没有匹配，则后面的 72 bit 就不会参与比较运算。

因此，每 80 bit 位中的低 8 bit 又可以称为预查找比特位，这个功能对于用户而言是透明的，但是需要用户精心安排数据结构，才能充分发挥这个特性，例如，把不同数据结构的标志号放在这低 8 bit。

从统计学上分析，如果所有数据足够随机化，每 256 个条目只会有一个条目匹配，也只有这个条目的后 72 bit 才会参与比较，这样消耗的功耗只有原来的 10%（=(256×8+72)/(256×80)）。

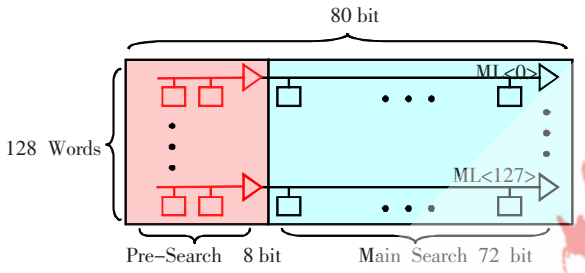


图 6 TCAM 预查找功能示意(IBM 32 nm)

值得一提的是，IBM 32 nm 工艺库的 TCAM 还同时提供功耗门控和深度休眠的方式来降低 TCAM 的动态功耗，前者对于使用者是透明的，后者需要设计相应的控制电路，而且从深度休眠的模式恢复到正常工作模式，至少需要 100 ns 的唤醒时间。

本文基于 IBM 32 nm 工艺库提供的 TCAM 和优先级编码器硬核，通过设计相应的外围电路，充分利用该 TCAM 的特性和硬核 IP，减少所需 TCAM 的块数和外围寄存器的数量，节省了 TCAM 在硅片上的摆放面积，同时降低了 TCAM 的功耗。本文提到的全部设计已经在公司的第 4 代以太网路由交换 ASIC 芯片上实现。

后续的工作，将研究如何基于厂家提供的 TCAM 如何进一步提高 TCAM 的查找性能。另一方面，将研究一些性能要求不高的场合下，如何充分利用 TCAM 的深度休眠功能，进一步降低整个芯片的功耗。

参考文献

[1] Embedded memory for Cu-32HP databook, SA15-6397-04, Revision 04[Z]. 2013.
[2] Embedded memory for Cu-45HP databook, SA15-6218-01, Revision 01[Z]. 2009.
[3] Huang Xiaohua. GAM cells and differential sense circuits for content addressable memory [P]. U.S: US6744653 B1,2004-06-01.
[4] ARSOVSKI I, WISTORT R. Self-referenced sense amplifier for across-chip-variation immune sensing in high performance Content-Addresable Memories[C]. Custom Integrated Circuits Conference, CICC'06, 2006:453-456.
[5] CHAO H J, LIU B. High performance switches and routers[M]. John Wiley & Sons, Inc., Publication, 2007.

（收稿日期：2013-06-05）

作者简介：

许俊，男，1972 年生，博士，主要研究方向：以太网交换路由芯片。

龚源泉，男，1980 年生，硕士，主要研究方向：以太网交换路由芯片。

李丽，女，1975 年生，副教授，主要研究方向：VSLI 设计技术和设计方法学。